



Virtual Browsing: A White Paper

Charlie Bennett & Seth Porter
Georgia Tech Library

Abstract

This paper proposes an addition to the Library's search and discovery services. In order to recreate one of the qualities of collection browsing that is lost by relocating the physical collection off-site, the Library should provide a recommendation service that can provide the kind of serendipitous search results that patrons have identified as critical-to-quality in their browsing. The paper details the proposed recommendation service and presents a SWOT analysis of creating the service.

Introduction

One particular quality of the Library's patrons' interactions with the collection, traditional shelf browsing, was radically changed when the Library moved the bulk of its physical collection to the Library Service Center.

Our patrons enhanced their personal discovery strategies by scanning the shelves while retrieving items from the stacks. They discovered unknown but interesting items that were, by coincidence and by classification, shelved near known items. While this discovery strategy has been identified by many of our patrons as an emotionally appealing and practical Library service, it is a by-product of open-stacks storage and the Library of Congress classification system. Traditional shelf browsing is impossible when the stacks are closed or the collection is off-site.

To recover this discovery strategy for patrons, we propose an addition to the Library's catalog services that creates a visual grid of item recommendations created by a recommendation engine that uses metadata and librarian input. This grid can be accessed by patrons if they want to browse and hidden if a patron is focused on known-item searches.

A virtual browsing system is vital to retain patrons and preserve their confidence in the Library. It has the additional benefit of keeping the physical and electronic collections visible.

The Library can provide patrons with many of the results of traditional shelf browsing through robust search and discovery tools. There are, however, serendipitous

discoveries that the Library cannot provide without creating a virtual browsing system for the catalog. Identifying virtual browsing as a specific search strategy for patrons will help ameliorate the resentment from more traditional users who feel that they have lost a valuable service. Younger patrons who consider recommendation engines to be integral parts of their consumer strategies will find a welcome familiarity to the virtual browsing system.

A virtual browsing system will have the advantage of not being limited to literal proximity on shelves. It will provide unknown but interesting items from any of the Library's (or its Interlibrary Loan partners') physical or electronic collections simultaneously.

This virtual browsing system does not replace the Library's catalog; it provides a set of search results modeled after our patrons' serendipitous discoveries in the open stacks while transcending the limits of the traditional browsing of physical shelves.

Environmental Scan

We scanned the library websites and web catalogs of our twenty peer institutions. Twelve have no virtual browsing component to their catalog search, six use Primo (or a very similar system), and two provide another service.

The twelve that have no virtual browsing component to their catalog search allow users to explore subject headings, other works by authors, and many other standard metadata-informed search extensions. The six with Primo-style virtual browsing allow users to "search the shelf" at the page for the item record, scrolling left or right through

their collections in call number order. The two peer institutions that provide another type of service are UCLA (<http://catalog.library.ucla.edu/vwebv/searchBasic>) and Virginia Tech (<https://addison.vt.edu>).

UCLA's library allows users to tag items and to collect items on personalized lists that are made public. This folksonomy is supported by information from WorldCat.

Virginia Tech's library, through their "Addison" services, allows users to create lists of books and supplements searches with information from LibraryThing, a social media site based around a public cataloging system.

The Georgia Tech Library has an opportunity to lead among its peers in creating a virtual browsing system that provides subject expert guidance, useful and community-based recommendations, and powerful serendipitous discovery.

Along with examining the online library catalogs of our peer institutions, we explored a pair of research engines, Yewno (<http://yewno.com/about/>) and The James Burke Institute's Knowledge Web (<http://k-web.org>). These tools are designed to help researchers explore online information, create search terms, and map knowledge concepts. Both tools are in development to some degree, and both tools provide assistance to researchers that is outside of the scope of our proposed virtual browsing service.

Overview

The following white paper is broken down into three sections 1.) The storyboard which is a visual representation of the virtual browsing outputs, including The Grid. 2.)

The strengths and weaknesses of virtual browsing systems, and the recommendation and design of the system proposed in the white paper 3.) The recommendation itself which details the recommendation engine, the three distinct frameworks, the details of how it would work, and what is needed for it to work.

The Story Board

Rather than attempt to describe through text we created a visual representation of the virtual browsing experience. The following narrative walks the reader through the virtual browsing experience from the time a user chooses a known item, labeled as an “anchor item,” to when the user chooses virtual browsing, to the results of the virtual browsing interface known as The Grid. The story board also gives the reader a brief introduction to the recommendation engine that drives virtual browsing—this will be covered in detail in the following sections. Please see the appendix for the complete virtual browsing narrative in a more visually appealing interface. Some visual sacrifices were made to embed the image in a word document.

Virtual Browsing

Georgia Tech Discovery Service



Virtual browsing initiates from a general discovery search, indexing all of Georgia Tech Libraries curated content and items accessible through ILL.

The user must choose an anchor item from results to initiate virtual browsing.

Behavioral Economics

© James Allison (James W.) 1983
Available at GTLSC General Collection HB74 .P8 A535 1983

Get It

Details

Virtual Browse



Personalize your results
[Edit](#)

SHOW ONLY
[Available in Print](#) (71)
[Full Text Online](#) (1,188)

REFINE MY RESULTS

Collection
[Conference Proceedings](#) (63)
[Core Collection](#) (6)
[More options](#)

Resource Type
[Books](#) (1,231)
[Journals](#) (9)
[Web Sites](#) (3)
[Databases](#) (1)
[Audio Visual](#) (1)
[More options](#)

Subject
[Business & Economics](#) (583)
[Health Care](#) (40)
[Social Sciences](#) (431)
[Psychiatry and Psychology](#) (404)
[Health Care Economics and Organizations](#) (417)
[More options](#)

Author/Creator
[SpringerLink \(Online service\)](#) (186)
[HathiTrust Inc.](#) (31)
[Council, National Research](#) (26)
[eLibrary, Inc.](#) (22)
[World Health Organization](#) (17)
[More options](#)

Creation Date
 From To [Refine](#)
 800 2017

Classification LOC

SHOW ONLY [Available in Print](#) (71) [Full Text Online](#) (1,188)

Behavioral economics
 James Allison (James W.) 1963
 Available at GTLSC General Collection HB74 P8 A535 1963



Virtual Browsing results i.e., THE GRID



Virtual Browsing Detailed Results



After the user chooses an anchor item from the search results and clicks the virtual browsing button the recommendation engine initiates three distinct results. The results are a browsable representation of three different recommendation frameworks. Each item will have one of three tags. The tags are the small images next to each of the three recommendation frameworks:

1.) Georgia Tech Recommendation Engine



Based on a unique recommendation engine, i.e., related items, citations in text, items that reference item, related searches, item-item connections and other practical connections.

2.) Library Connections

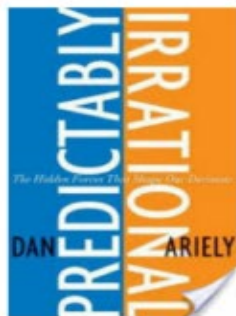


Recommendations based on record meta-data, author, subject, subject headings, and more.

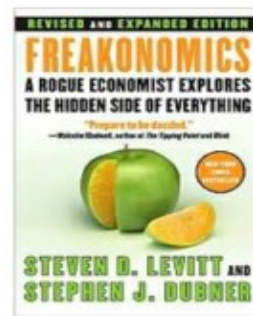
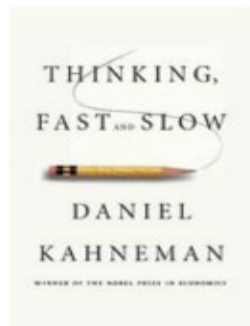
3.) Patron Connections



Recommendations based on Librarian/Archivist/Faculty, and curriculum based suggestions mined from the LMS that creates a course cross walk between course meta-data and library and archival resources e.g., required readings in foundational courses.



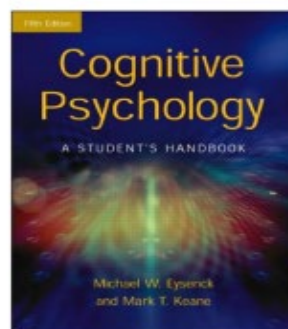
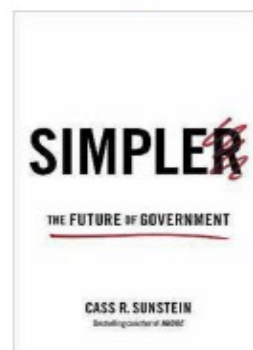
Georgia Tech Library



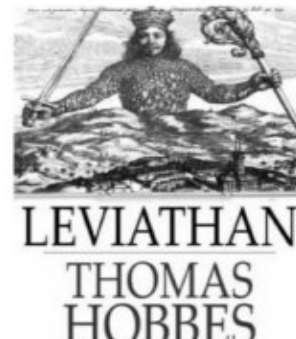
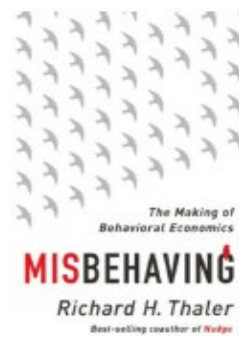
Georgia Tech Library



Georgia Tech Library



Georgia Tech Library



Strengths & Weaknesses

The following section covers the potential strengths and weaknesses of the recommended virtual browsing system. The strengths include accessibility, holistic discovery through the library portal, usability, serendipitous discovery, aesthetic design/wow factor, and leadership in digital content access. The weaknesses include privacy, scale, user comprehension, and technology requirements. The analysis should assist the library in understanding the importance of virtual browsing and specifically, this recommendation, but it will also describe potential problems that must be overcome.

Strengths

Accessibility

Virtual browsing initiates from a known item search, and will assist the user in accessing needed content. Currently, the virtual browsing system in Primo gives the user access to material within its LOC taxonomy. While this can assist a user in accessing related content, it is also very narrow and does not replicate the accessibility of the physical library. The virtual browsing recommendation engine will alleviate this problem by assisting the user in accessing all related content. It will do so through three different frameworks

illustrated in the story board section. The removal of physical content from the Georgia Tech library has created considerable uncertainty for many of our patrons who are more comfortable browsing the stacks in a traditional library. While we cannot recreate this completely, we can recreate the accessibility of the content through the virtual browsing interface known as The Grid. The Grid will be a vital part of the 21st century library and will alleviate many of the access concerns of our user base.

Holistic discovery through library portal

The virtual browsing recommendation engine will create holistic discovery through all the curated content available in the Georgia Tech library website and accessible through ILL. It will do so through a unique recommendation engine with three different outputs.

The Georgia Tech Recommendation Engine (GTRE)

One potential output once a user chooses an anchor item is the Georgia Tech Recommendation Engine. The GTRE is an algorithm that suggests a related content based on the item selected. The algorithm mines related items, items other users accessed who viewed the anchor item, citations within the text (much like web of science), related searches, and item-item connections (Mayer-Schonberger & Cukier

2014). Essentially, an Item-item recommendation system match each of the user's purchased, viewed, and rated items to similar items, then combines those similar items into a recommendation list (Linden, Smith, and York, 2003).

Library Connections

The library connections will be related to the meta-data within a traditional item record. When a user chooses an anchor item the library connections will suggest related items based on the anchor item's meta-data, related authors, subject headings, keywords, and the LOC classification.

Patron Connections

The patron connections will be more of an institutionally intimate framework based on the subject and institutional knowledge of a subject librarian or archivist, faculty, and required and suggested readings mined from the institutional LMS. When a user chooses an anchor item The Grid will visualize related items based on the input of the librarian, archivist, faculty, and curriculum based suggestions by creating a course cross walk between course meta-data and library and archival resources, e.g., required readings in foundational courses (Yaco et. al 2016). This will create a discovery service based on the deep knowledge of the Georgia Tech community.

Usability

The virtual browsing interface named The Grid will assist users in locating and accessing needed content with ease. The browse-ability of The Grid is ideal for the modern user. It is modeled after the aesthetically pleasing interfaces of commercial products like Amazon, Hulu, and Netflix. The Grid with its visual representation of the material is minimal, intuitive, and most importantly easy to use. The design should avoid clutter and provide clear, unambiguous results for a user. The Grid will display needed content based on the three different frameworks. Each framework represents a different information need for the user and this is clear with visual representation called tags under each item. Much like commercial products like Primo the user will choose an anchor item from the results and be able to access the record or “preview” of the material. The intuitive browse-ability of the curated content through The Grid will create an enjoyable experience that replicates the physical experience of browsing in the stacks. While curated content accessibility is a vital aspect of the 21st century digital library, the ease of access is just as important. A modern user will turn away from a digital interface if the design is not intuitive and easy to use. While the recommendation engine will provide the user with related content access, the user needs to be able to find the information easily, and intuitively. The Grid will provide this.

Serendipitous discovery

Content access through browsing the physical stacks of a library has always been an engaging and serendipitous experience. Users could get lost finding connections between their known item search and their greater research question by spending time physically searching and browsing the stacks. While the current catalog and digital library can replicate the known item search and certain aspects of the browsing experience, they cannot provide the serendipitous search results that the user now misses with offsite storage. However, the unique recommendation engine using the three frameworks, and the visual appealing results through The Grid will help ameliorate the traditional browsing experience and result in serendipitous discovery. When a user searches a known item and chooses to virtually browse, they will be able to access content that is related to their search question based on the unique recommendation engine, item record and meta-data, subject specific recommendation by librarian/archivist input, and required and suggested readings mined from the LMS. It will feel almost as if the user is in the stacks, choosing items, seeing the connections, and even asking the librarian for their suggestions based on the previous browsing experience. The serendipitous results will widen the content access, but they

will also be tailored to the users search and they will help connect lines of inquiry that a traditional LOC taxonomy would not find.

Aesthetic design/wow factor

Visual design is an important part of the human computer interaction. It even connects deeply to the learning experience. Aesthetic design positively impacts the usability of items and digital interfaces (Kurosu & Kashimura, 1995; Glore 2011; Boulton 2005; Hancock 2004). Not only is aesthetic design vital in creating a “wow” factor around a specific tool, it is important in usability of that tool. Currently, the virtual browsing tool is a simple visual interface of the books near the item recalled and in the LOC range. Not only is this not ideal for accessibility and item recall, it is far from ideal from a design standpoint. The Grid will fix this concern. It will follow the best practices in graphic design and user experience in online interactions. The results will be visual representations of the item, but it will also be minimal, intuitive, coherent, and consistent, with clear tags on each item. The Grid will be easy to browse. Additionally, it be consistent with the best practices in commercial applications and implement A/B testing of the design effects on each application. A/B testing is “a method of comparing two versions of a webpage or app against each other to determine which one performs better.

AB testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal (Optimizely 2017).” This process will help improve the Grid and the user experience. The Grid will not only have “wow factor” it will be beautiful, and beautiful design assists in usability (Tractinsky, 2014).

The combination of these three frameworks will provide a holistic overview for access to a research subject and will make connections to items that would have otherwise not been accessed through the library portal.

Leadership in digital content access

The virtual browsing system is based on the unique recommendation engine that takes into account the three disparate but related frameworks that will put the Georgia Tech library as a the leader in digital content access in academic libraries. As previously described peer institutions like UCLA, Virginia Tech, and more have limited virtual browsing capabilities, however, none, will provide the access, holistic and serendipitous discovery, as the Georgia Tech virtual browsing tool. The virtual browsing tool will not replace the catalog, it will be an addition to the portfolio of tools for the user. Additionally, the combination of the three pronged recommendation engine,

the attractiveness and usability of The Grid, and the access to curated content will make this tool a leader not only among our peers, but commercial enterprises as well.

Weaknesses

Privacy

As codified by the ALA, user privacy is a core value of librarianship. Users require privacy and confidentiality to exercise their intellectual freedom (ALA 2004). Any recommendation engine that tracks, saves, and makes use of searching behavior to inform its results will create some concerns in patrons regarding their privacy and thus reduce their trust in the library.

Some researchers may resist the idea that the connections they make in their research might be reflected in public results: for instance, innovation in interdisciplinary fields can start with recognizing the similarities in two very disparate works, making that connection part of the researcher's intellectual property in some way.

The virtual browsing system must acknowledge these privacy concerns and clearly explain the way the system uses user behavior. User should be able to opt-out from having their browsing behavior recorded and used for the virtual

browsing system. This opt-out system has several benefits: clearly addressing user privacy concerns; increasing trust in the library and the virtual browsing system; and making clear the benefits of using the Library's planned personalized account and mobile app.

Scale

The richer and more thorough the metadata attached to an item, the more the recommendation engine will be able to do when a patron chooses an anchor item to create the recommendation grid. Some items, however, may not have rich enough metadata to produce a robust set of recommendations. In the most extreme cases, metadata-poor items could be clustered in a particular research area, making the recommendation engine perform poorly for a specific group of library patrons.

The recommendation engine uses automatic processes for many of its recommendations but the affinities that most reflect the community of users come from librarian input; subject matter experts identify individual connections and affinities for items to enrich the grid. This action cannot be scaled up to the millions of items in the catalog, and the hundreds of millions of items available to the library patron from the world of information.

Patron Comprehension

Patrons have not been taught to expect recommendations from the library catalog, and often recommendations are signifiers of a commercial service. The interface of the virtual browsing recommendation system, including the web buttons, images, and layout of the Grid, must help the patrons recognize the purpose and rationale to the recommendations the Library provides. The design and explanatory text must provide context and explanation for 1) the metaphorical connection to physical browsing, 2) the codification of serendipity as a service, and 3) the use of metadata, LMS data, user behavior, and librarian/archivist knowledge for recommendations. The design will require a significant amount of work and testing to be sure patrons can easily comprehend the virtual browsing system.

Technology Requirements

The virtual browsing system will require the Library to build or buy technology that can 1) harvest data from Georgia Tech's LMS to automate the patron recommendations based on syllabi, 2) interact with WorldCat, Amazon, and other bibliographic data, and 3) can process the metadata and full text of items. This system will also interact with servers the Library does not own. This technology will require significant time, effort, knowledge, and resources to develop.

Virtual Browsing Recommendation

In this section, we provide details of the recommendation engine and its three distinct frameworks. We examine how the engine would work and the requirements for its operation, including several options for how the engine can be developed.

The Framework

The virtual browsing system cannot create a knowledge constellation of all possible resources. However, with careful design the virtual browsing system can replicate and improve on the traditional experience of browsing the stacks in the physical library. The virtual browsing system will need to contain all of the library's curated content as well as any content that is accessible through interlibrary loan. To do so the Georgia Tech Library virtual browsing service should follow the framework and recommendation described previously and detailed below. Furthermore, the display will need to be visual representation of the content, and should be very similar to The Grid described previously, so the user has access to related items, but also has a pleasant search experience. Additionally, each of the three frameworks will need to be labeled and described clearly so that the user has complete transparency on why the content is being recommended based on their initial known item search. The output of the virtual browsing project should follow the subsequent framework.

The Georgia Tech Recommendation Engine (GTRE)

The first output once a user chooses an anchor item is the Georgia Tech Recommendation Engine. The *GTRE* is an algorithm that suggests a related content based on the item selected. This algorithm is similar to commercial products like Amazon, Netflix, and Hulu. This will be based on data analytics which as mentioned previously is based on the “item-item” algorithm from Amazon. Item-Item searching filters and matches each of the user’s purchased and rated items to similar items, then combines those similar items into a recommendation list. The *GTRE* would need to track the users data and combine it with an algorithm that also incorporates citations within the text (much like web of science), and items citing the anchor item. One thing the *GTRE* will need to do besides the “item-item” connections, and the citation analysis is cull old data so the recommendations will be accurate and up to date. The *GTRE* is one aspect of the virtual browsing recommendation. As previously explored this will be one of the three visual tags on The Grid.

Library Connections

The *Library Connections* will be related to the meta-data within a traditional item record. When a user chooses an anchor item the library connections will suggest related items based on the anchor item’s meta-data, related authors, subject headings, keywords, and the LOC classification. The *Library Connections* recommendation is based on the more traditional role of the library. The *Library*

Connections output will connect the user with relevant items based on the meta-data of the item. Ideally, this will be an improved version of traditional browsing by call number. As previously explored, this will be one of the three visual tags on The Grid.

Patron Connections

The *Patron Connections* will be more of an intimate framework based on the subject and institutional knowledge of a subject librarian or archivist, faculty, and required and suggested readings mined from the institutional LMS. When a user chooses an anchor item The Grid will visualize related items based on the input of the librarian, archivist, faculty, and curriculum based suggestions e.g., required readings in foundational courses. This will create a discovery service based on the deep knowledge and needs of the Georgia Tech community. For this to work successfully the Georgia Tech library will need access to all the required and suggested reading of the institution. We will need to data mine these requirements and institutional LMS to create recommendations and create a tailored course cross walk between course meta-data and library and archival resource (Yaco et. al 2016). Additionally, faculty, Librarians, and Archivist would be able to feed related content into this framework as suggested readings for a specific subject. In creating this specific framework for the recommendation engine the Georgia Tech library will help make the curriculum of the institution intimately involved in the recommendation of the library.

Browser Plugin

In addition to the virtual browsing system, we recommend a browser plugin that mines item meta-data and is added into the virtual browsing content. Additionally, it can be automatically accessible through inter library loan. A similar project is the open access button. It finds requested material and integrates it with library catalogs and ILL system to find accessible copies through a library discovery system, and fulfills the ILL requests instantly (OA Button 2017). The plugin will 1) be an extension of the current “Find it @ GT” Service, 2) be connected to the users Georgia Tech account, and 3) be usable in WorldCat and Google Scholar. When a user is on either of these sites and they find an item that they wish to see more of, they will click on the plugin and that will import the item record and meta-data into the virtual browsing framework. This will be a user-friendly process of improving item recall and accessibility of content for the user. The browser plugin will be a one-of-a-kind digital library tool that sets the Georgia Tech library apart and makes physical content digitally accessible.

Development

The development of the virtual browsing system is an element of the technology portfolio in library next. However, the virtual browsing white paper has the developed three potential options that could assist their decision making process.

Option One: The priority for development should be placed on the framework to adopt the three recommendation engines that make up the virtual browsing service. The Georgia Tech Library should hire an outside consultant or developer to create the recommendation engine and The Grid.

Option Two: The Georgia Tech Library creates a hackathon or prize to help develop the framework for the algorithm and The Grid. Netflix uses a similar model, called the Netflix Prize, to improve their recommendation engine (Netflix Prize 2009).

Option Three: The Georgia Tech library partners with a graduate level course in computer science or machine learning to develop the recommendation engines and The Grid. This could potentially be a graduate level thesis for an individual student or a group project to be developed throughout the semester or year.

The Plugin: The plugin is an important part of the discovery aspect of the virtual browsing system. The development of the plugin is vital, however, depending on the resource allocation it should be of a second priority to the virtual browsing recommendation framework and the grid.

The development strategy should follow the steps detailed above. Additionally, electronic resources and digital library experts will need to help develop a system for item record and meta-data import and curation that is interoperable with the Georgia Tech library's catalog. Additionally there is potential partners to develop this idea with such as Ex Libris, Alma/Primo, OA Button, or other institutions with Ex Libris open API and developers network.

The plugin's development will be an extension of the current "Find it @GT" and not an entirely new service. The developers will need to have a thorough knowledge of the current system as they develop the add on.

Conclusion

Providing our patrons with a virtual replacement for the experience of browsing the shelf will require substantial developer and designer work. This work is necessary to create a virtual browsing system that can accomplish two critical tasks: successfully deliver serendipitous search results in the online catalog to mimic serendipitous discovery of items in the stacks; and use as much of the available data about the Library's items and the Georgia Tech community's needs to make those serendipitous recommendations more useful than the results shelf browsing could produce.

This virtual browsing system will fulfill a need that is very important to a small but vocal group of our patrons, and that will bolster the Library's reputation on campus. Additionally, the data mining aspect has a sustaining usefulness in helping the library meet current research needs and to identify new trends with GT research interests. Ultimately, this kind of tool will make the library and archival resources more relevant for its primary users.

References

- (2017). "A/B Testing." Optimization Gallery. Retrieved from:
<https://www.optimizely.com/ab-testing/>
- (2017). "Yewno." Retrieved from: <https://about.yewno.com/>
- (2017). "James Burke Knowledge Web." Retrieved from: (2017). "Yewno." Retrieved from: <https://about.yewno.com/>
- (2017). "Open Access Button." Retrieved from:
<https://blog.openaccessbutton.org/putting-the-oa-into-interlibrary-loan-9d5fa5eb25d4>
- (2009). "Netflix Prize." Retrieved from: <http://www.netflixprize.com/>
- (2004). "Core Values of Librarianship." Retrieved from:
<http://www.ala.org/advocacy/intfreedom/statementspols/corevalues>
- Boulton, M. (2005). "Aesthetic-Usability Effect." Retrieved from:
<http://www.markboulton.co.uk/journal/aesthetic-usability-effect>
- Hancock, D. (2004). Improving the Environment in Distance Learning Courses through the Application of Aesthetic Principles.
http://etd.fcla.edu/SF/SFE0000426/Thesis_Project_Hancock.pdf
- Kurosu, M., & Kashimura, K. (1995). Apparent usability vs inherent usability: Experimental analysis on the determinants of the apparent usability.
- Glore, P. (2011). The Impact of Design and Aesthetics on Usability, Credibility, and Learning in an Online Environment.
http://www.westga.edu/~distance/ojdla/winter134/david_glore134.html
- Linden, G. Smith, B. York, J. (2003). "Item-to-Item Collaborative Filtering." Amazon.com Recommendations. An Industry Report. Retrieved from:
<https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>
- Mayer-Schonberger, V. Cukier, K. (2014). "Big Data: A Revolution That Will Transform How We Live, Work, and Think." Eamon Dolan/Mariner Books

Tractinsky, Noam (2014): Visual Aesthetics. In: Soegaard, Mads and Dam, Rikke Friis (eds.). *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*. Aarhus, Denmark: The Interaction Design Foundation. Available online at https://www.interaction-design.org/encyclopedia/visual_aesthetics.html

Yaco, S. Brown, C. Konrad, L (2016). "Linking Special Collections to Classrooms: A Curriculum-to-Collection Crosswalk." *The American Archivist Vol. 79, No. 2 Fall/Winter 2016*